PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/54, 9/10, 1/21 | A1

(11) International Publication Number: **WO 00/14249**

(43) International Publication Date: 16 March 2000 (16.03.00)

(21) International Application Number: PCT/EP98/05573

(22) International Filing Date: 2 September 1998 (02.09.98)

(71) Applicant *(for all designated States except US)*: PLANTTEC BIOTECHNOLOGIE GMBH [DE/DE]; Forschung & Entwicklung, Hermannswerder 14, D–14473 Potsdam (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: QUANZ, Martin [DE/DE]; Oppelner Strasse 34, D–10997 Berlin (DE). PROVART, Nicholas [CA/DE]; Parforceheide 31, D–14163 Berlin (DE).

(74) Agent: VOSSIUS & PARTNER; P.O. Box 86 07 67, D–81634 München (DE).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

(54) Title: NUCLEIC ACID MOLECULES ENCODING AN AMYLOSUCRASE

(57) Abstract

Described are nucleic acid molecules which encode an amylosucrase as well as methods for the production of α–1,4 glucans and fructose using such nucleic acid molecules or the encoded proteins. Furthermore, described are host cells transformed with the described nucleic acid molecules.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Nucleic acid molecules encoding an amylosucrase

The present invention relates to nucleic acid molecules encoding a protein having amylosucrase activity and to vectors containing such molecules. Furthermore, the invention relates to the production of α-1,4 glucans and fructose using the described nucleic acid molecules or the encoded proteins.

Linear α-1,4 glucans are polysaccharides consisting of glucose monomers, the latter being exclusively linked to each other by α-1,4 glycosidic bonds. The most frequently occurring natural α-1,4 glucan is amylose, a component of plant starch. Recently, more and more importance has been attached to the commercial use of linear α-1,4 glucans. Due to its physico-chemical properties amylose can be used to produce films that are colorless, odorless and flavorless, non-toxic and biologically degradable. Already today, there are various possibilities of application, e.g., in the food industry, the textile industry, the glass fiber industry and in the production of paper.

One has also succeeded in producing fibers from amylose whose properties are similar to those of natural cellulose fibers and which allow to partially or even completely replace them in the production of paper. Being the most important representative of the linear α-1,4 glucans, amylose is particularly used as binder for the production of tablets, as thickener of puddings and creams, as gelatin substitute, as binder in the production of sound-insulating wall panels and to improve the flow properties of waxy oils. Another property of the α-1,4 glucans, which recently has gained increasing attention, is the capability of these molecules to form inclusion compounds with organic complexers due to their helical structure. This property allows to use the α-1,4 glucans for a wide variety of applications. Present considerations relate to their use for the molecular encapsulation of vitamins, pharmaceutical compounds and aromatic substances, as well as their use for the chromatographic separation of mixtures of substances over immobilized linear α-1,4

glucans. Amylose also serves as starting material for the production of so-called cyclodextrins (also referred to as cycloamyloses, cyclomaltoses) which in turn are widely used in the pharmaceutical industry, food processing technology, cosmetic industry and analytic separation technology. These cyclodextrins are cyclic maltooligosaccharides from 6-8 monosaccharide units, which are freely soluble in water but have a hydrophobic cavity which can be utilized to form inclusion compounds.

Today, α-1,4 glucans, in particular linear α-1,4 glucans, are obtained in the form of amylose from starch. Starch itself consists of two components. One component forms the amylose as an unbranched chain of α-1,4 linked glucose units. The other component forms the amylopectin, a highly branched polymer from glucose units in which in addition to the α-1,4 links the glucose chains can also be branched via α-1,6 links. Due to their different structure and the resulting physico-chemical properties, the two components are also used for different fields of application. In order to be able to directly utilize the properties of the individual components, it is necessary to obtain them in pure form. Both components can be obtained from starch, the process, however, requiring several purification steps and involving considerable time and money. Therefore, there is a need to find possibilities of obtaining both components of the starch in a uniform manner. It is known that certain bacteria, in particular those of the genus Neisseria produce enzymes capable of synthesizing linear α-1,4 glucans from sucrose. In order to be able to use such enzymes for the efficient production of α-1,4 glucans, it is necessary to isolate and characterize the corresponding DNA sequences.

The technical problem underlying the present invention is therefore to provide nucleic acid molecules and processes that allow the production of α-1,4 glucans.

The solution of this technical problem is achieved by the present invention by providing the embodiments characterized in the claims.

The invention therefore relates to nucleic acid molecules encoding a protein having the enzymatic activity of an amylosucrase selected from the group consisting of

(a) nucleic acid molecules encoding a protein comprising the amino acid sequence as depicted in SEQ ID NO: 2;

(b) nucleic acid molecules comprising the nucleotide sequence of the coding region as indicated in SEQ ID NO: 1;

(c) nucleic acid molecules encoding an analogue of the polypeptide having the amino acid sequence as depicted under SEQ ID NO: 2; and

(d) nucleic acid molecules, the sequence of which differs from the sequence of a nucleic acid molecule as defined in (c) due to the degeneracy of the genetic code.

The nucleic acid sequence of the coding region depicted in SEQ ID NO: 1 encodes a protein of Neisseria polysaccharea having the enzymatic activity of an amylosucrase. With the help of the nucleic acid molecules of the present invention it is possible to produce microorganisms and fungi, particularly yeasts, that are capable of producing an enzyme catalyzing the synthesis of α-1,4 glucans from sucrose.

It is furthermore possible to produce at low production costs α-1,4 glucans, in particular linear α-1,4 glucans, as well as pure fructose syrup with the help of the DNA sequences of the invention or of the proteins encoded by them.

Nucleotide sequences which encodes an analogue of the polypeptide as depicted in SEQ ID NO: 2 are understood in the scope of the present invention as nucleotide sequence which encode a polypeptide having the following characteristics:

(a) it has amylosucrase activity; and preferably,

(b) it furthermore shows an identity on the amino acid sequence level of at least 80%, more preferably of at least 85%, even more preferably of at least 90% and particularly preferred of at least 95%, to the amino acid sequence as depicted in SEQ ID NO: 2 over its complete length.

Thus, the present invention also relates to nucleic acid molecules encoding a polypeptide the sequence of which differs at one or more positions from the amino

acid sequence as depicted in SEQ ID NO: 2 and which still has amylosucrase activity. The differences in the amino acid sequence may be due to replacements of amino acid residues by other amino acid residues, to the addition of amino acid residues, preferably at the N- or C-terminus of the polypeptide, or to deletions of on or more amino acid residues, preferably at the N- or C-terminus of the protein. The generation of nucleic acid molecules encoding such analogues of the described protein is well within the common general knowledge of the person skilled in the art.

The present invention also relates to nucleic acid molecules the complementary strand of which hybridizes under stringent conditions to a nucleic acid molecule as defined above and which encode a polypeptide having the enzymatic activity of an amylosucrase.

In this invention the term "hybridization" means a hybridization under stringent conditions as described for example in Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). "Stringent conditions" mean that there is a sequence identity of at least 80% of the complete coding sequence, preferably an identity of at least 90%, more preferably of at least 95% and particularly preferred of at least 99%.

Nucleic acid molecules hybridizing to the molecules according to the invention may be isolated e.g. from genomic or from cDNA libraries produced from organism expressing an amylosucrase, for example, from microorganisms, in particular from bacteria of the genus Neisseria. The identification and isolation of such nucleic acid molecules may take place by using the molecules according to the invention or parts of these molecules or, as the case may be, the reverse complement strands of these molecules, e.g. by hybridization according to standard methods (see e.g. Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

As a probe for hybridization e.g. nucleic acid molecules may be used which exactly or basically contain the nucleotide sequence of the coding region indicated under SEQ ID NO. 1 or parts thereof. The fragments used as hybridization probe may also be synthetic fragments which were produced by means of the conventional synthesizing

methods and the sequence of which is basically identical with that of a nucleic acid molecule according to the invention. After identifying and isolating the genes hybridizing to the nucleic acid sequences according to the invention, the sequence has to be determined and the properties of the proteins encoded by this sequence have to be analyzed.

The molecules hybridizing to the nucleic acid molecules of the invention also comprise fragments, derivatives and allelic variants of the above-described nucleic acid molecules which encode a protein having the enzymatic activity of an amylosucrase. Thereby, fragments are defined as parts of the nucleic acid molecules, which are long enough in order to encode a protein still having the enzymatic activity. This includes also parts of nucleic acid molecules according to the invention which lack the nucleotide sequence encoding the signal peptide responsible for the secretion of the protein. The term derivatives means that the sequences of these molecules differ from the sequences of the above-mentioned nucleic acid molecules at one or more positions and that they exhibit a high degree of homology to these sequences. Hereby, homology means a sequence identity of at least 80%, in particular an identity of at least 90%, preferably of more than 95% and still more preferably a sequence identity of more than 98%. The deviations occurring when comparing with the above-described nucleic acid molecules might have been caused by deletion, substitution, insertion or recombination.

Moreover, homology means that functional and/or structural equivalence exists between the respective nucleic acid molecules or the proteins they encode. The nucleic acid molecules, which are homologous to the above-described molecules and represent derivatives of these molecules, are generally variations of these molecules, that constitute modifications which exert the same biological function. These variations may be naturally occurring variations, for example sequences derived from other organisms, or mutations, whereby these mutations may have occurred naturally or they may have been introduced by means of a specific mutagenesis. Moreover the variations may be synthetically produced sequences. The allelic variants may be naturally occurring as well as synthetically produced variants or variants produced by recombinant DNA techniques.

The proteins encoded by the various variants of the nucleic acid molecules according to the invention exhibit certain common characteristics. Enzyme activity, molecular weight, immunologic reactivity, conformation etc. may belong to these characteristics as well as physical properties such as the mobility in gel electrophoresis, chromatographic characteristics, sedimentation coefficients, solubility, spectroscopic properties, stability, pH-optimum, temperature-optimum etc.

An amylosucrase (also referred to as sucrose:1,4-α glucan 4-α-glucosyltransferase, E.C. 2.4.1.4.) is an enzyme for which the following reaction scheme is suggested:

sucrose + $(\alpha\text{-}1,4\text{-D-glucosyl})_n \rightarrow$ D-fructose + $(\alpha\text{-}1,4\text{-D-glucosyl})_{n+1}$

This reaction is a transglucosylation. The transglucosylation can take place in the presence or absence of acceptor molecules. Such acceptor molecules can be polysaccharides, such as maltooligosaccharides, dextrin, glycogen etc. When such an acceptor molecule is a linear, oligomeric α-1,4-glucan, the resulting product is a polymeric linear α-1,4-glucan. When the transglucosylation catalyzed by the amylosucrase is carried out in the absence of such acceptor molecule, a glucan is obtained which comprises a terminal fructose molecule. All the products obtainable by transglycosylation with the help of an amylosucrase in the absence or presence of an acceptor molecule are referred to in the scope of the present invention as α-1,4 glucans.

The reaction mechanism for a transglucosylation by an amylosucrase in the absence of an acceptor molecule can be described as follows:

$G\text{-}F+n(G\text{-}F) \rightarrow G_n\text{-}G\text{-}F+nF$,

wherein G-F is sucrose, G is glucose, F is fructose and G_n-G-F is an α-1,4 glucan.

The reaction mechanism in the presence of an acceptor molecule can be described as follows:

$mG\text{-}F+G_n \rightarrow G_{n+m}+mF$,

wherein G_n is a polysaccharide acceptor molecule, G_{n+m} is the polysaccharide plus α-1,4 glucan chains added thereto by amylosucrase, G-F is sucrose, F is fructose and G is glucose. The products of the reaction catalyzed by an amylosucrase are the

above described α-1,4 glucans and fructose. Cofactors are not required. Amylosucrase activity so far has been found only in few bacteria species, among them particularly the species *Neisseria* (MacKenzie et al., Can. J. Microbiol. 24 (1978), 357-362) and the enzyme has been examined only for its enzymatic activity. According to Okada et al., the partially purified enzyme from *Neisseria perflava* upon addition of sucrose results in the synthesis of glycogen-like polysaccharides which are branched to a small extent (Okada et al., J. Biol. Chem. 249 (1974), 126-135). Likewise, the intra- or extracellularly synthesized glucans of *Neisseria perflava* and *Neisseria polysaccharea* exhibit a certain degree of branching (Riou et al., Can. J. Microbiol. 32 (1986), 909-911). Whether these branches are introduced by the amylosucrase or via another enzyme that is present in the purified amylosucrase preparations as contamination, has so far not been elucidated. Since an enzyme introducing branching has so far not been found, it is assumed that both the polymerization and the branching reactions are catalyzed by amylosucrase (Okada et al., loc. cit.).

The enzyme that is expressed in a constitutive manner in *Neisseria* is extremely stable, binds very strongly to the polymerization products and is competitively inhibited by the product fructose (MacKenzie et al., Can. J. Microbiol. 23 (1977), 1303-1307). The Neisseria species *Neisseria polysaccharea* secretes the amylosucrase (Riou et al., loc. cit.) while in the other *Neisseria* species it remains in the cell. Enzymes having amylosucrase activity could only be detected in microorganisms. Plants are not known to have amylosucrases.

The detection of the enzymatic activity of the amylosucrase can be achieved by detecting the synthesized glucans, as is described in Example 3, below. Detection is usually carried out by using a iodine stain. It is possible to identify bacterial colonies expressing amylosucrase by, e.g., treatment with iodine vapor. Colonies synthesizing the α-1,4 glucans are stained blue.

The enzyme activity of the purified enzyme can be detected on, e.g., sucrose-containing agarose plates. If the protein is applied to such a plate and incubated for about 1 h or more at 37°C, it diffuses into the agarose and catalyzes the synthesis of glucans. The latter can be detected by treatment with iodine vapor. Furthermore, the

protein can be detected in native polyacrylamide gels. After a native polyacrylamide gel electrophoresis, the gel is equilibrated in sodium citrate buffer (50 mM, pH 6.5) and incubated over night in a sucrose solution (5% in sodium citrate buffer). If the gel is subsequently stained with Lugol's solution, areas in which proteins having amylosucrase activity are localized are stained blue due to the synthesis of α-1,4 glucans.

The protein encoded by a nucleic acid molecule according to the invention preferably has a molecular weight of 63± 20kDa, more preferably of 63± 15kDa and even more preferably of 63± 10kDa when determined in an SDS-PAGE.

In a preferred embodiment, the invention relates to nucleic acid molecules encoding an amylosucrase from a microorganism, particularly a gram negative microorganism, preferably from a bacterium of the species *Neisseria* and particularly preferred from *Neisseria polysaccharea*.

The nucleic acid molecules according to the invention can be any kind if nucleic acid molecule, for example, RNA or DNA, in particular cDNA or genomic DNA. They can be synthetic, partly synthetic or isolated from natural sources.

Furthermore, the present invention relates to vectors, for example, plasmids, phages, cosmids, phagemids or artificial chromosomes, containing a nucleic acid molecule according to the invention. The invention particularly relates to vectors in which the nucleic acid molecule of the invention is linked to sequences ensuring expression of the nucleic acid molecule in prokaryotic or eukaryotic host cells. Expression in this regard means transcription, preferably transcription and translation. Expression vectors have been extensively described in the art. In addition to a selection marker gene and a replication origin allowing replication in the selected host they normally contain a promoter active in the host cell and a transcription termination signal. Between promoter and termination signal there is normally at least one restriction site or one polylinker which allows insertion of a coding DNA sequence. As promoter sequence the DNA sequence which normally controls transcription of the corresponding gene can be used as long as it is active in the selected organism. This sequence can be replaced by other promoter sequences. Promoters can be used

that effect constitutive expression of the gene or inducible promoters that allow a selective regulation of the expression of the gene downstream thereof. Bacterial and viral promoter sequences for the expression in prokaryotic host cells have been extensively described in the art. Promoters allowing a particularly strong expression of the gene downstream thereof, are, e.g., the T7 promoter (Studier et al., in Methods in Enzymology 185 (1990), 60-89), lacuv5, trp, trp-lacUV5 (DeBoer et al., in Rodriguez, R.L. and Chamberlin, M.J., (Eds.), Promoters, Structure and Function; Praeger, New York, 1982, pp. 462-481; DeBoer et al., Proc. Natl. Acad. Sci. USA 80 (1983), 21-25), lp_1, rac (Boros et al., Gene 42 (1986), 97-100) or the ompF promoter. Vectors for the expression of heterologous genes in yeasts have also been described (e.g., Bitter et al., Methods in Enzymology 153 (1987), 516-544). These vectors, in addition to a selection marker gene and a replication origin for the propagation in bacteria, contain at least one further selection marker gene that allows identification of transformed yeast cells, a DNA sequence allowing replication in yeasts and a polylinker for the insertion of the desired expression cassette. The expression cassette is constructed from promoter, DNA sequence to be expressed and a DNA sequence allowing transcriptional termination and polyadenylation of the transcript. Promoters and transcriptional termination signals from *Saccharomyces* have also been described and are available. An expression vector can be introduced into yeast cells by transformation according to standard techniques (Methods in Yeast Genetics, A Laboratory Course Manual, Cold Spring Harbor Laboratory Press, 1990). Cells containing the vector are selected and propagated on appropriate selection media. Yeasts furthermore allow to integrate the expression cassette via homologous recombination into the genome of a cell using an appropriate vector, leading to a stable inheritance of the feature.

Furthermore, the present invention relates to host cells transformed with a nucleic acid molecule or with a vector according to the invention. Suitable host cells are prokaryotic cells, such as microorganisms, e.g. bacteria, such as E. coli, Bacillus, Streptococcus etc., or eukaryotic cells, e.g. fungal cells, such as Saccharomyces cerevisiae; plant cells or animal cells, e.g. insect cells, CHO cells etc.

Moreover, the present invention relates to a process for producing a protein having amylosucrase activity comprising culturing a host cell according to the invention under conditions allowing expression of the protein and recovering the protein from the cells and/or the culture medium.

The present invention also relates to a protein having the enzymatic activity of an amylosucrase which is encoded by a nucleic acid molecule according to the invention, or which is obtainable by the process according to the invention.

In another aspect the present invention relates to a process for producing α-1,4 glucans and/or fructose comprising

(a) culturing a host cell according to the invention which secrets the amylosucrase into the culture medium in a medium which contains sucrose and under conditions allowing expression and secretion of the amylosucrase; and

(b) recovering the produced α-1,4 glucans and/or the fructose from the culture medium.

The above described process now allows to produce pure α-1,4 glucans in vitro. The amylosucrase expressed by *Neisseria polysaccharea* is an extracellular enzyme which synthesizes linear α-1,4 glucans outside of the cells on the basis of sucrose. Unlike in the most pathways of synthesis for polysaccharides that proceed within the cell, neither activated glucose derivatives nor cofactors are required. The energy that is required for the formation of the α-1,4 glucosidic link between the condensed glucose residues is directly obtained from the hydrolysis of the link between the glucose and the fructose unit in the sucrose molecule.

It is therefore possible to cultivate amylosucrase-secreting host cells in a sucrose-containing medium, with the secreted amylosucrase leading to a synthesis of α-1,4 glucans from sucrose in the medium. These glucans can be isolated from the culture medium.

Furthermore, the process according to the invention allows to produce in an inexpensive manner pure fructose syrup. Conventional methods for the production of fructose either contemplate the enzymatic hydrolysis of sucrose using an invertase or the degradation of starch into glucose units, often by acidolysis, and subsequent enzymatic conversion of the glucose into fructose by glucose isomerase. Both methods result in mixtures of glucose and fructose. The two components have to be separated from each other by chromatographic processes which are time consuming and expensive.

In the process according of the invention, the separation of the substrate, sucrose, from the two reaction products, fructose and α-1,4 glucans, or separation of the two reaction products can be achieved by, e.g., using membranes allowing the permeation of fructose but not of sucrose or glucans. If the fructose is continuously removed via such a membrane, the sucrose is converted more or less completely into fructose and linear glucans.

Also the amylosucrase producing cells can preferably be immobilized on a carrier material located between two membranes, one of which allows the permeation of fructose but not of sucrose or glucans and the other allows the permeation of sucrose but not of glucans. The substrate is supplied through the membrane which allows the permeation of sucrose. The synthesized glucans remain in the space between the two membranes and the released fructose can continuously be removed from the reaction equilibrium through the membrane which allows only the permeation of fructose. Such a set-up allows an efficient separation of the reaction products and thus inter alia the production of pure fructose.

The use of amylosucrases for the production of pure fructose offers the advantage that the comparably inexpensive substrate sucrose can be used as starting material and furthermore that the fructose can be isolated from the reaction mixture in a simple manner without chromatographic processes.

In a preferred embodiment the host cells used in the process is a microorganism, such as Saccharomyces cerevisiae or E. coli, and even more preferably the host cell is immobilized. Immobilization generally is achieved by inclusion of the cells in an

appropriate material such as, e.g., alginate, polyacrylamide, gelatin, cellulose or chitosan. It is, however, also possible to adsorb or covalently bind the cells to a carrier material (Brodelius and Mosbach, in Methods in Enzymology, Vol. 135:173-175). An advantage of the immobilization of cells is that considerably higher cell densities can be achieved than by cultivation in a liquid culture, resulting in a higher productivity. Also the costs for agitation and ventilation of the culture as well as for the measures for maintaining sterility are reduced. An important aspect is that immobilization allows a continuous production so that long unproductive phases which usually occur in fermentation processes can be avoided or can at least be considerably reduced. As mentioned above, yeast cells expressing an amylosucrase can be used as a microorganism in the process. Cultivation methods for yeasts have been sufficiently described (Methods in Yeast Genetics, A Laboratory Course Manual, Cold Spring Harbor Laboratory Press, 1990). Immobilization of the yeasts is also possible and is already used in the commercial production of ethanol (Nagashima et al., in Methods in Enzymology 136, 394-405; Nojima and Yamada, in Methods in Enzymology 136, 380-394).

However, the use of yeasts secreting amylosucrase for the synthesis of α-1,4 glucans in sucrose-containing media is not readily possible as yeasts secrete an invertase that hydrolyzes extracellular sucrose. The yeasts import the resulting hexoses via a hexose transporter. Gozalbo and Hohmann (Current Genetics 17 (1990), 77-79), however, describe a yeast strain that carries a defective suc2 gene and that therefore cannot secrete invertase. Also, these yeast cells do not contain a transport system for importing sucrose into the cells. If such a strain is modified with the nucleic acid molecule of the invention such that it secretes an amylosucrase into the culture medium, α-1,4 glucans are synthesized by the amylosucrase if the culture medium contains sucrose. The fructose being formed as reaction product may subsequently be imported by the yeasts.

Furthermore, the present invention relates to a process for the production of α-1,4 glucans and/or fructose in vitro comprising the step of bringing a protein according to the invention into contact with a sucrose-containing solution under conditions

allowing the conversion of sucrose to α-1,4 glucans and fructose and recovering the produced α-1,4 glucans and/or fructose from the solution.

In particular, it is possible to synthesize α-1,4 glucans *in vitro* with the help of a cell-free enzyme preparation. This may be obtained, for example, by cultivating amylosucrase-secreting host cells in a sucrose-free medium allowing expression of the amylosucrase until the stationary growth phase is reached. After removal of the cells from the growth medium by centrifugation the secreted enzyme can be obtained from the supernatant. The enzyme can then be added to sucrose-containing solutions to synthesize α-1,4 glucans and fructose. As compared to the synthesis of α-1,4 glucans directly in a sucrose-containing growth medium this method is advantageous in that the reaction conditions can be better controlled and that the reaction products are substantially purer and can more easily be further purified.

The enzyme can be purified from the culture medium by conventional purification techniques such as precipitation, ion exchange chromatography, affinity chromatography, gel filtration, HPLC reverse phase chromatography, etc.

It is furthermore possible to express a polypeptide by modification of the DNA sequence inserted into the expression vector leading to a polypeptide which can be isolated more easily from the culture medium due to certain properties. It is possible to express the enzyme as a fusion protein along with another polypeptide sequence whose specific binding properties allow isolation of the fusion protein via affinity chromatography.

Known techniques are, e.g., expression as fusion protein along with glutathion S transferase and subsequent purification via affinity chromatography on a glutathion column, making use of the affinity of the glutathion S transferase to glutathion (Smith and Johnson, Gene 67 (1988), 31-40). Another known technique is the expression as fusion protein along with the *maltose binding protein* (MBP) and subsequent purification on an amylose column (Guan et al., Gene 67 (1988), 21-30; Maina et al., Gene 74 (1988), 365-373).

In a preferred embodiment, the amylosucrase in such a process is immobilized.

In addition to the possibility of directly adding the purified enzyme to a sucrose-containing solution to synthesize α-1,4 glucans, there is the alternative of

immobilizing the enzyme on a carrier material. Such immobilization offers the advantage that the enzyme as synthesis catalyst can easily be retrieved and can be used several times. Since the purification of enzymes usually is very time and cost intensive, an immobilization and reuse of the enzyme contributes to a considerable reduction of the costs. Another advantage is the high degree of purity of the reaction products which inter alia is due to the fact that the reaction conditions can be better controlled when immobilized enzymes are used. The insoluble linear glucans yielded as reaction products can then be easily purified further.
There are many carrier materials available for the immobilization of proteins which can be coupled to the carrier material either by covalent or non-covalent links (for an overview see: Methods in Enzymology Vol. 135, 136 and 137). Widely used carrier materials are, e.g., agarose, cellulose, polyacrylamide, silica or nylon.

A further possibility of the use of proteins having amylosucrase activity is to use them for the production of cyclodextrins. Cyclodextrins are produced by the degradation of starch by the enzyme cyclodextrin transglycosylase (EC 2.4.1.19) which is obtained from the bacterium *Bacillus macerans*. Due to the branching of starch only about 40% of the glucose units can be converted to cyclodextrins using this system. By providing substantially pure proteins having amylosucrase activity it is possible to synthesize cyclodextrins on the basis of sucrose under the simultaneous action of amylosucrase and cyclodextrin transglycosylase, with the amylosucrase catalyzing the synthesis of linear glucans from sucrose and the cyclodextrin transglycosylase catalyzing the conversion of these glucans into cyclodextrins.

Abbreviations used

IPTG	isopropyl ß-D-thiogalacto-pyranoside

Media and solutions used

YT medium	8 g bacto-tryptone

	5 g yeast extract 5 g NaCl ad 1000 ml with ddH_2O
YT plates	YT medium with 15 g bacto-agar/ 1000 ml
Lugol's solution	12 g KI 6 g I_2 ad 1.8 l with ddH_2O

The examples serve to illustrate the invention.

Example 1

Isolation of a genomic DNA sequence coding for an amylosucrase activity from *Neisseria polysaccharea*

For the isolation of a DNA sequence coding for an amylosucrase activity from *Neisseria polysaccharea* first a genomic DNA library was established. *Neisseria polysaccharea* cells were cultured on "Columbia blood agar" (Difco) for 2 days at 37°C. The resulting colonies were harvested from the plates. Genomic DNA was isolated according to the method of Ausubel et al. (in: Current Protocols in Molecular Biology (1987), J. Wiley & Sons, NY) and processed. The DNA thus obtained was partially digested with the restriction endonuclease *Sau3A*. The resulting DNA fragments were ligated into the *BamH*I digested vector pBluescript SK(-). The ligation products were transformed in *E. coli* XL1-Blue cells. For their selection, the cells were plated onto YT plates with ampicillin as selection marker. The selection medium additionally contained 5% sucrose and 1 mM IPTG. After incubation over night at 37°C the bacterial colonies that had formed were stained with iodine by placing crystalline iodine into the lid of a petri dish and placing the culture dishes with the

bacteria colonies for 10 min each conversely onto the petri dish. The iodine which evaporated at room temperature stained some regions of the culture dishes that contained amylose-like glucans blue. From bacteria colonies that showed a blue corona plasmid DNA was isolated according to the method of Birnboim & Doly (Nucleic Acids Res. 7 (1979), 1513-1523). Said DNA was retransformed in *E. coli* SURE cells. The transformed cells were plated onto YT plates with ampicillin as selection marker. Positive clones were isolated.

Example 2

Sequence analysis of the genomic DNA insert of the plasmid pNB2

From an *E. coli* clone obtained according to working example 1 a recombinant plasmid was isolated. Restriction analyses showed that said plasmid was a ligation product consisting of two vector molecules and an approx. 4.2 kb long genomic fragment. The plasmid was digested with the restriction endonuclease *Pst*I and the genomic fragment was isolated (GeneClean, Bio101). The fragment thus obtained was ligated into a pBluescript II SK vector linearized with *Pst*I, resulting in a duplication of the *Pst*I and *Sma*I restriction sites. The ligation product was transformed in *E. coli* cells and the latter were plated on ampicillin plates for selection. Positive clones were isolated. From one of these clones a plasmid was isolated and part of the sequence of its genomic DNA insert was determined by standard techniques using the dideoxy method (Sanger et al., Proc. Natl. Acad. Sci. USA 74 (1977), 5463-5467). The entire insert is approx. 4.2 kbp long. The nucleotide sequence was determined and is indicated in SEQ ID NO. 1.

Example 3

Expression of an extracellular amylosucrase activity in transformed *E. coli* cells

(a) Detection of an amylosucrase activity during growth on YT plates

For the expression of an extracellular amylosucrase activity, *E. coli* cells were transformed with the isolated plasmid vector according to standard techniques. A colony of the transformed strain was incubated on YT plates (1.5% agar; 100 µg/ml ampicillin; 5% sucrose; 0.2 mM IPTG) over night at 37°C. The amylosucrase activity was detected by subjecting the colonies to iodine vapor. Amylosucrase-expressing colonies exhibit a blue corona. Amylosucrase activity can be observed even if no IPTG was present, probably due to the activity of the endogenous amylosucrase promoters.

(b) Detection of an amylosucrase activity during growth in YT medium

For the expression of an extracellular amylosucrase activity, *E. coli* were transformed with the isolated plasmid vector according to standard techniques. YT medium (100 µg/ml ampicillin; 5% sucrose) was inoculated with a colony of the transformed strain. The cells were incubated over night at 37°C under constant agitation (rotation mixer; 150-200 rpm). The products of the reaction catalyzed by amylosucrase were detected by adding Lugol's solution to the culture supernatant, leading to blue staining.

(c) Detection of the amylosucrase activity in the culture supernatants of transformed *E. coli* cells which were cultivated without sucrose

For the expression of an extracellular amylosucrase activity, *E. coli* cells were transformed with the isolated plasmid vector according to standard techniques. YT medium (100 µg/ml ampicillin) was inoculated with a colony of the transformed strain. The cells were incubated over night at 37°C under constant agitation (rotation mixer; 150-200 rpm). Then the cells were removed by centrifugation (30 min, 4°C, 5500 rpm, JA10 Beckmann rotor). The supernatant was filtered through a 0.2 µm filter (Schleicher & Schuell) under sterile conditions.
Detection of an amylosucrase activity was carried out by

(i) incubating the supernatant on a sucrose-containing agar plate. 40 µl of the supernatant were placed in a whole punched into an agar plate (5% sucrose in 50 mM sodium citrate buffer pH 6.5) and incubated at least for one hour at 37°C. The products of the reaction catalyzed by amylosucrase were detected by staining with iodine vapor. Presence of the reaction products produces a blue stain.

(ii) or by gel electrophoretic separation of the proteins of the supernatant in a native gel and detection of the reaction products in the gel after incubation with sucrose. 40-80µl of the supernatant were separated by gel electrophoresis on an 8% native polyacrylamide gel (0.375 M Tris pH 8.8) at a voltage of 100 V. The gel was then twice equilibrated 15 min with approx. 100 ml 50 mM sodium citrate buffer (pH 6.5) and incubated over night at 37°C in sodium citrate buffer pH 6.5/5% sucrose. In order to make the reaction product of the reaction catalyzed by amylosucrase visible, the gel was rinsed with Lugol's solution. Bands having amylosucrase activity were stained blue.

Example 4

In vitro production of glucans with partially purified amylosucrase

For the expression of an extracellular amylosucrase activity, *E. coli* cells were transformed with the isolated plasmid vector according to standard techniques. YT medium (100 µg/ml ampicillin) was inoculated with a colony of the transformed strain. The cells were incubated over night at 37°C under constant agitation (rotation mixer; 150-200 rpm). Then the cells were removed by centrifugation (30 min, 4°C, 5500 rpm, JA10 Beckmann rotor). The supernatant was filtered through a 0.2 µm filter (Schleicher & Schuell) under sterile conditions.

The supernatant was then concentrated by 200 times using an Amicon chamber (YM30 membrane having an exclusion size of 30 kDa, company Amicon) under pressure (p=3 bar). The concentrated supernatant was added to 50 ml of a sucrose

solution (5% sucrose in 50 mM sodium citrate buffer pH 6.5). The entire solution was incubated at 37°C. Whitish insoluble polysaccharides are formed.

Example 5

Characterization of the reaction products synthesized by amylosucrase from Example 4

The insoluble reaction products described in Example 4 are soluble in 1 M NaOH. The reaction products were characterized by measuring the absorption maximum. Approx. 100 mg of the isolated reaction products (wet weight) were dissolved in 200 µl 1 M NaOH and diluted with H_2O 1:10. 900 µl of 0.1 M NaOH and 1 ml Lugol's solution were added to 100 µl of this dilution. The absorption spectrum was measured between 400 and 700 nm. The maximum is 605 nm (absorption maximum of amylose: approx. 614 nm).

HPLC analysis of the reaction mixture of Example 4 on a CARBOPAC PA1 column (DIONEX) showed that in addition to the insoluble products soluble products were also formed. These soluble products are short-chained polysaccharides. The chain length was between approx. 5 and approx. 60 glucose units. To a smaller extent, however, even shorter or longer molecules could be detected.

With the available analytical methods it was not possible to detect branching in the synthesis products.

Example 6

Expression of an intracellular amylosucrase activity in transformed *E. coli* cells

Using a polymerase chain reaction (PCR) a fragment was amplified from the isolated plasmid vector which comprises the nucleotides 981 to 2871 of the sequence depicted in SEQ ID NO. 1. The following oligonucleotides were used as primers:

TPN2 5' - CTC ACC ATG GGC ATC TTG GAC ATC - 3'
(SEQ ID NO. 3)

TPC1 5' - CTG CCA TGG TTC AGA CGG CAT TTG G - 3'
(SEQ ID NO. 4)

The resulting fragment contains the coding region for amylosucrase except for the nucleotides coding for the 16 N-terminal amino acids. These amino acids comprise the sequences that appear to be necessary for the secretion of the enzyme from the cell. Furthermore, this PCR fragment contains 88 bp of the 3' untranslated region. By way of the primers used *Nco*I restriction sites were introduced into both ends of the fragment.

After digestion with the restriction endonuclease *Nco*I the resulting fragment was ligated with the *Nco*I digested expression vector pMex 7. The ligation products were transformed in *E. coli* cells and transformed clones were selected. Positive clones were incubated over night at 37°C on YT plates (1.5% agar; 100 µg/ml ampicillin; 5% sucrose; 0.2 mM IPTG). After subjecting the plates to iodine vapor no blue staining could be observed in the area surrounding the bacteria colonies, but the intracellular production of glycogen could be detected (brown staining of transformed cells in contrast to no staining in nontransformed XL1-Blue cells). In order to examine the functionality of the protein, transformed cells cultivated on YT medium were broken up by ultrasound and the obtained crude extract was pipetted onto sucrose-containing agar plates. After subjecting the plates to iodine vapor a blue stain could be observed.

CLAIMS

1. A nucleic acid molecule encoding a protein having the enzymatic activity of an amylosucrase, selected from the group consisting of
   (a) nucleic acid molecules encoding a protein comprising the amino acid sequence depicted under SEQ ID NO. 2;
   (b) nucleic acid molecules comprising the coding region depicted under SEQ ID NO. 1;
   (c) nucleic acid molecules encoding an analogue of the polypeptide having the amino acid sequence as depicted under SEQ ID NO: 2; and
   (d) nucleic acid molecules the sequence of which differs from the sequence of a nucleic acid molecule as defined in (c) due to the degeneracy of the genetic code.

2. The nucleic acid molecule of claim 1 which is genomic DNA.

3. A vector containing a nucleic acid molecule of claim 1 or 2.

4. The vector of claim 3, in which the nucleic acid molecule encoding a protein having the enzymatic activity of an amylosucrase is functionally linked to sequences allowing expression in prokaryotic or eukaryotic host cells.

5. A host cell transformed with a nucleic acid molecule of claim 1 or 2 or with a vector of claim 3 or 4.

6. A process for producing a protein with the enzymatic activity of an amylosucrase comprising culturing the host cell of claim 5 under conditions allowing expression of the amylosucrase and recovering the protein from the cells and/or the culture medium.

7. A protein having the enzymatic activity of an amylosucrase which is encoded by a nucleic acid molecule of claim 1 or 2 or which is obtainable by the process of claim 6.

8. A process for the production of α-1,4 glucans and/or fructose comprising
   (a) culturing a host cell of claim 5 which secrets the amylosucrase into the culture medium in a sucrose-containing culture medium under conditions allowing expression and secretion of the amylosucrase; and
   (b) recovering the produced α-1,4 glucans and/or fructose from the culture medium.

9. The process of claim 8, wherein the host cell is a microorganism.

10. The process of claim 9 or 10, wherein the host cell is immobilized.

11. A process for the production of α-1,4 glucans and/or fructose in vitro comprising
   (a) contacting a sucrose-containing solution with a protein of claim 7 under conditions allowing the conversion of sucrose to α-1,4 glucans and fructose by the amylosucrase; and
   (b) recovering the produced α-1,4 glucans and/or fructose from the solution.

12. The process of claim 11, wherein the protein is immobilized.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: PlantTec Biotechnologie GmbH Forschung & Entwicklung
(B) STREET: Hermannswerder 14
(C) CITY: Potsdam
(E) COUNTRY: Germany
(F) POSTAL CODE (ZIP): 14473

(ii) TITLE OF INVENTION: Nucleic acid molecules encoding an amylosucrase

(iii) NUMBER OF SEQUENCES: 4

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4173 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Neisseria polysaccharea

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1971..3878

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GATCGCCTTC GCCCAATTGC GACCAAAGTT TTTTGGTAAA CAGCTTGGGG TTGTTCTCGA 60

TGACTTTGTT GGCGATTTTG AGAATCCGCG CGGTGGAGCG GTAGTTTTGC TCCAGTTTGA 120

TGACCTTCAT CTGCGGATAG TTTTCCTGCA TTTTGCGCAG GTTTTCCATG TTCGCGCCGC 180

GCCATGCGTA GATGGACTGG TCGTCGTCGC CGACGGCGGT AAACATCCCT TCCGCGCCGG 240

TCAGAAGTTT CATCAACGTA AATTGGCAGG TATTCGTATC TTGGCATTCG TCAACCAGCA 300

GATAACGCAG CCGCCGCTGC CATTTGTTGC GCACTTCGCT GTTTTGCTGC AACAGCACGG 360

CAGGCAGGCG GATTAAGTCA TCGAAGTCCA CTGCCTGATA GCTTTGTAAG GTTTCCTGAT 420

AGCTCGCATA CACGCGTGCG GTTTGTTGTT CCCAAATGTT CGATGCCGTC TGAACGACAT 480

CTTCAGGCGT TTTTAAATCG TTTTTCCAAA GGGAAATTTG ATGTTGCGCT TTGAATGTGG 540

CTTCTTTGCC CGTACCGCCT AAGAGTTCGC CGATGATTTT CGCGCTGTCG GTAGAGTCGA 600

GGATGGAGAA GTTTTTTTTG TAACCGATAT GGTTCGCCTC TTCGCGCAAA ATCTTCATGC 660

CCAAAGAATG GAACGTGCAA ATTGTCAGCC CGCGCGTTTG CGATTTGGGC AGCATTTTGG 720

CGACGCGCTC CTGCATTTCC GCAGCGGCTT TGTTGGTAAA GGTAATCGCG GCGACGGTAT 780

GCGGCAGATA GCCGACATTG ACAATCAAAT GCTTGATTTT TTGAGTAATC ACGCCGGTTT 840

TTCCGCTGCC CGCGCCTGCA AGGACGAGCA GGGGGCCGCC GAGGTAGCGG ACGGCTTCGA 900

GCTGTTGGGG ATTGAGTTTC ATCATGTTTT GATGCCGTCT GAAATCAGTC TGCGCCGCTT 960

TCGAGGCAGT CGAGTGCCGC ACGGAGGGCG GATACGCCGA TTTGCCCCGG CGCGGAGTTT 1020

TGCGTTCCCG AACCGAACGT GATGCTTGAG CCGAACACCT GTCCGGCAAG GCGGCTGACC 1080

GCCCCCTTTT GCCCCATCGA CATCGTAACA ATCGGTTTGG TGGCAAGCTC TTTCGCTTTG 1140

AGCGTGGCAG AAAGCAAAGT CAGCACGTCT TCCGCGCTTT GCGGCATCAC CGCAATTTTG 1200

CAGATGTCCG CGCCGCAGTC CTCCATCTGT TTCAGACGGC ATACGATTTC TTCTTGCGGC 1260

GGCGTGCGGT GAAACTCATG ATTGCAGAGC AGGGCGGCGA TGCCGTTTTT TTGAGCATGC 1320

GCCACGGCGC GCCGGACGGC GGTTTCGCCG GAAAAAAGCT CGATATCGAT AATGTCGGGC 1380

AGGCGGCTTT CAATCAGCGA GTCGAGCAGT TCAAAATAAT AATCGTCCGA ACACGGGAAC 1440

GAGCCGCCTT CGCCATGCCG TCTGAACGTA AACAGCAGCG GCTTGTCGGG CAGCGCGTCG 1500

CGGACGGTCT GCGTGTGGCG CAATACTTCG CCGATGCTGC CCGCGCATTC CAAAAAATCG 1560

GCGCGGAACT CGACGATATC GAAGGGCAGG TTTTTGATTT GGTCAAGTAC GGCGGAAAGT 1620

```
ACGGCGGCAT CGCGGGCGAC AAGCGGCACG GCGATTTTGG TGCGTCCGCT TCCGATAACG     1680

GTGTTTTTGA CGGTCAGGCT GGTGTGCATG GCGGTTGTTG CGGCTGAAAG GAACGGTAAA     1740

GACGCAATTA TAGCAAAGGC ACAGGCAATG TTTCAGACGG CATTTCTGTG CGGCCGGCTT     1800

GATATGAATC AAGCAGCATC CGCATATCGG AATGCAGACT TGGCACAAGC CCTGTCTTTT     1860

CTAGTCAGTC CGCAGTTCTT GCAGTATGAT TGCACGACAC GCCCTACACG GCATTTGCAG     1920

GATACGGCGG CAGACCGCCG GTCGGAAACT TCAGAATCGG AGCAGGCATC ATG TTG       1976
                                                          Met Leu
                                                            1

ACC CCC ACG CAG CAA GTC GGT TTG ATT TTA CAG TAC CTC AAA ACA CGC       2024
Thr Pro Thr Gln Gln Val Gly Leu Ile Leu Gln Tyr Leu Lys Thr Arg
          5                  10                  15

ATC TTG GAC ATC TAC ACG CCC GAA CAG CGC GCC GGC ATC GAA AAA TCC       2072
Ile Leu Asp Ile Tyr Thr Pro Glu Gln Arg Ala Gly Ile Glu Lys Ser
     20                  25                  30

GAA GAC TGG CGG CAG TTT TCG CGC CGC ATG GAT ACG CAT TTC CCC AAA       2120
Glu Asp Trp Arg Gln Phe Ser Arg Arg Met Asp Thr His Phe Pro Lys
 35                  40                  45                  50

CTG ATG AAC GAA CTC GAC AGC GTG TAC GGC AAC AAC GAA GCC CTG CTG       2168
Leu Met Asn Glu Leu Asp Ser Val Tyr Gly Asn Asn Glu Ala Leu Leu
                 55                  60                  65

CCT ATG CTG GAA ATG CTG CTG GCG CAG GCA TGG CAA AGC TAT TCC CAA       2216
Pro Met Leu Glu Met Leu Leu Ala Gln Ala Trp Gln Ser Tyr Ser Gln
            70                  75                  80

CGC AAC TCA TCC TTA AAA GAT ATC GAT ATC GCG CGC GAA AAC AAC CCC       2264
Arg Asn Ser Ser Leu Lys Asp Ile Asp Ile Ala Arg Glu Asn Asn Pro
          85                  90                  95

GAT TGG ATT TTG TCC AAC AAA CAA GTC GGC GGC GTG TGC TAC GTT GAT       2312
Asp Trp Ile Leu Ser Asn Lys Gln Val Gly Gly Val Cys Tyr Val Asp
     100                 105                 110

TTG TTT GCC GGC GAT TTG AAG GGC TTG AAA GAT AAA ATT CCT TAT TTT       2360
Leu Phe Ala Gly Asp Leu Lys Gly Leu Lys Asp Lys Ile Pro Tyr Phe
115                 120                 125                 130

CAA GAG CTT GGT TTG ACT TAT CTG CAC CTG ATG CCG CTG TTT AAA TGC       2408
Gln Glu Leu Gly Leu Thr Tyr Leu His Leu Met Pro Leu Phe Lys Cys
                135                 140                 145
```

```
CCT GAA GGC AAA AGC GAC GGC GGC TAT GCG GTC AGC AGC TAC CGC GAT      2456
Pro Glu Gly Lys Ser Asp Gly Gly Tyr Ala Val Ser Ser Tyr Arg Asp
            150                 155                 160

GTC AAT CCG GCA CTG GGC ACA ATA GGC GAC TTG CGC GAA GTC ATT GCT      2504
Val Asn Pro Ala Leu Gly Thr Ile Gly Asp Leu Arg Glu Val Ile Ala
        165                 170                 175

GCG CTG CAC GAA GCC GGC ATT TCC GCC GTC GTC GAT TTT ATC TTC AAC      2552
Ala Leu His Glu Ala Gly Ile Ser Ala Val Val Asp Phe Ile Phe Asn
    180                 185                 190

CAC ACC TCC AAC GAA CAC GAA TGG GCG CAA CGC TGC GCC GCC GGC GAC      2600
His Thr Ser Asn Glu His Glu Trp Ala Gln Arg Cys Ala Ala Gly Asp
195                 200                 205                 210

CCG CTT TTC GAC AAT TTC TAC TAT ATT TTC CCC GAC CGC CGG ATG CCC      2648
Pro Leu Phe Asp Asn Phe Tyr Tyr Ile Phe Pro Asp Arg Arg Met Pro
                215                 220                 225

GAC CAA TAC GAC CGC ACC CTG CGC GAA ATC TTC CCC GAC CAG CAC CCG      2696
Asp Gln Tyr Asp Arg Thr Leu Arg Glu Ile Phe Pro Asp Gln His Pro
            230                 235                 240

GGC GGC TTC TCG CAA CTG GAA GAC GGA CGC TGG GTG TGG ACG ACC TTC      2744
Gly Gly Phe Ser Gln Leu Glu Asp Gly Arg Trp Val Trp Thr Thr Phe
        245                 250                 255

AAT TCC TTC CAA TGG GAC TTG AAT TAC AGC AAC CCG TGG GTA TTC CGC      2792
Asn Ser Phe Gln Trp Asp Leu Asn Tyr Ser Asn Pro Trp Val Phe Arg
    260                 265                 270

GCA ATG GCG GGC GAA ATG CTG TTC CTT GCC AAC TTG GGC GTT GAC ATC      2840
Ala Met Ala Gly Glu Met Leu Phe Leu Ala Asn Leu Gly Val Asp Ile
275                 280                 285                 290

CTG CGT ATG GAT GCG GTT GCC TTT ATT TGG AAA CAA ATG GGG ACA AGC      2888
Leu Arg Met Asp Ala Val Ala Phe Ile Trp Lys Gln Met Gly Thr Ser
                295                 300                 305

TGC GAA AAC CTG CCG CAG GCG CAC GCC CTC ATC CGC GCG TTC AAT GCC      2936
Cys Glu Asn Leu Pro Gln Ala His Ala Leu Ile Arg Ala Phe Asn Ala
            310                 315                 320

GTT ATG CGT ATT GCC GCG CCC GCC GTG TTC TTC AAA TCC GAA GCC ATC      2984
Val Met Arg Ile Ala Ala Pro Ala Val Phe Phe Lys Ser Glu Ala Ile
        325                 330                 335

GTC CAC CCC GAC CAA GTC GTC CAA TAC ATC GGG CAG GAC GAA TGC CAA      3032
Val His Pro Asp Gln Val Val Gln Tyr Ile Gly Gln Asp Glu Cys Gln
    340                 345                 350
```

```
ATC GGT TAC AAC CCC CTG CAA ATG GCA TTG TTG TGG AAC ACC CTT GCC      3080
Ile Gly Tyr Asn Pro Leu Gln Met Ala Leu Leu Trp Asn Thr Leu Ala
355                 360                 365                 370

ACG CGC GAA GTC AAC CTG CTC CAT CAG GCG CTG ACC TAC CGC CAC AAC      3128
Thr Arg Glu Val Asn Leu Leu His Gln Ala Leu Thr Tyr Arg His Asn
                375                 380                 385

CTG CCC GAG CAT ACC GCC TGG GTC AAC TAC GTC CGC AGC CAC GAC GAC      3176
Leu Pro Glu His Thr Ala Trp Val Asn Tyr Val Arg Ser His Asp Asp
            390                 395                 400

ATC GGC TGG ACG TTT GCC GAT GAA GAC GCG GCA TAT CTG GGC ATA AGC      3224
Ile Gly Trp Thr Phe Ala Asp Glu Asp Ala Ala Tyr Leu Gly Ile Ser
        405                 410                 415

GGC TAC GAC CAC CGC CAA TTC CTC AAC CGC TTC TTC GTC AAC CGT TTC      3272
Gly Tyr Asp His Arg Gln Phe Leu Asn Arg Phe Phe Val Asn Arg Phe
    420                 425                 430

GAC GGC AGC TTC GCT CGT GGC GTA CCG TTC CAA TAC AAC CCA AGC ACA      3320
Asp Gly Ser Phe Ala Arg Gly Val Pro Phe Gln Tyr Asn Pro Ser Thr
435                 440                 445                 450

GGC GAC TGC CGT GTC AGT GGT ACA GCC GCG GCA TTG GTC GGC TTG GCG      3368
Gly Asp Cys Arg Val Ser Gly Thr Ala Ala Ala Leu Val Gly Leu Ala
                455                 460                 465

CAA GAC GAT CCC CAC GCC GTT GAC CGC ATC AAA CTC TTG TAC AGC ATT      3416
Gln Asp Asp Pro His Ala Val Asp Arg Ile Lys Leu Leu Tyr Ser Ile
            470                 475                 480

GCT TTG AGT ACC GGC GGT CTG CCG CTG ATT TAC CTA GGC GAC GAA GTG      3464
Ala Leu Ser Thr Gly Gly Leu Pro Leu Ile Tyr Leu Gly Asp Glu Val
        485                 490                 495

GGT ACG CTC AAT GAC GAC GAC TGG TCG CAA GAC AGC AAT AAG AGC GAC      3512
Gly Thr Leu Asn Asp Asp Asp Trp Ser Gln Asp Ser Asn Lys Ser Asp
    500                 505                 510

GAC AGC CGT TGG GCG CAC CGT CCG CGC TAC AAC GAA GCC CTG TAC GCG      3560
Asp Ser Arg Trp Ala His Arg Pro Arg Tyr Asn Glu Ala Leu Tyr Ala
515                 520                 525                 530

CAA CGC AAC GAT CCG TCG ACC GCA GCC GGG CAA ATC TAT CAG GGC TTG      3608
Gln Arg Asn Asp Pro Ser Thr Ala Ala Gly Gln Ile Tyr Gln Gly Leu
                535                 540                 545
```

```
CGC CAT ATG ATT GCC GTC CGC CAA AGC AAT CCG CGC TTC GAC GGC GGC      3656
Arg His Met Ile Ala Val Arg Gln Ser Asn Pro Arg Phe Asp Gly Gly
            550                 555                 560

AGG CTG GTT ACA TTC AAC ACC AAC AAC AAG CAC ATC ATC GGC TAC ATC      3704
Arg Leu Val Thr Phe Asn Thr Asn Asn Lys His Ile Ile Gly Tyr Ile
        565                 570                 575

CGC AAC AAT GCG CTT TTG GCA TTC GGT AAC TTC AGC GAA TAT CCG CAA      3752
Arg Asn Asn Ala Leu Leu Ala Phe Gly Asn Phe Ser Glu Tyr Pro Gln
    580                 585                 590

ACC GTT ACC GCG CAT ACC CTG CAA GCC ATG CCC TTC AAG GCG CAC GAC      3800
Thr Val Thr Ala His Thr Leu Gln Ala Met Pro Phe Lys Ala His Asp
595                 600                 605                 610

CTC ATC GGT GGC AAA ACT GTC AGC CTG AAT CAG GAT TTG ACG CTT CAG      3848
Leu Ile Gly Gly Lys Thr Val Ser Leu Asn Gln Asp Leu Thr Leu Gln
                615                 620                 625

CCC TAT CAG GTC ATG TGG CTC GAA ATC GCC TGACGCACGC TTCCCAAATG        3898
Pro Tyr Gln Val Met Trp Leu Glu Ile Ala
            630                 635

CCGTCTGAAC CGTTTCAGAC GGCATTTGCG CCGAAGCGGA CGGTAGTCCC CAAAAGGGAA    3958

ACATGCGATA ATAGCCGCCC ATCACATCCC GCGCCGCAGC CCGTGTTGCG CCGCATCCCA    4018

CATACCGCAT TTGTTCCGGA GTAACCCCAA TGTCAGACGA CAAAAGCAAA GCCCTTGCCG    4078

CCGCACTGGC GCAAATCGAA AAAAGTTTCG GCAAAGGCGC CATCATGAAA ATGGACGGCA    4138

GCCAGCAGGA AGAAAACCTC GAAGTCATTT CCACC                               4173
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 636 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Leu Thr Pro Thr Gln Gln Val Gly Leu Ile Leu Gln Tyr Leu Lys
  1               5                  10                  15

Thr Arg Ile Leu Asp Ile Tyr Thr Pro Glu Gln Arg Ala Gly Ile Glu
            20                  25                  30
```

Lys Ser Glu Asp Trp Arg Gln Phe Ser Arg Arg Met Asp Thr His Phe
35 40 45

Pro Lys Leu Met Asn Glu Leu Asp Ser Val Tyr Gly Asn Asn Glu Ala
50 55 60

Leu Leu Pro Met Leu Glu Met Leu Leu Ala Gln Ala Trp Gln Ser Tyr
65 70 75 80

Ser Gln Arg Asn Ser Ser Leu Lys Asp Ile Asp Ile Ala Arg Glu Asn
85 90 95

Asn Pro Asp Trp Ile Leu Ser Asn Lys Gln Val Gly Gly Val Cys Tyr
100 105 110

Val Asp Leu Phe Ala Gly Asp Leu Lys Gly Leu Lys Asp Lys Ile Pro
115 120 125

Tyr Phe Gln Glu Leu Gly Leu Thr Tyr Leu His Leu Met Pro Leu Phe
130 135 140

Lys Cys Pro Glu Gly Lys Ser Asp Gly Gly Tyr Ala Val Ser Ser Tyr
145 150 155 160

Arg Asp Val Asn Pro Ala Leu Gly Thr Ile Gly Asp Leu Arg Glu Val
165 170 175

Ile Ala Ala Leu His Glu Ala Gly Ile Ser Ala Val Val Asp Phe Ile
180 185 190

Phe Asn His Thr Ser Asn Glu His Glu Trp Ala Gln Arg Cys Ala Ala
195 200 205

Gly Asp Pro Leu Phe Asp Asn Phe Tyr Tyr Ile Phe Pro Asp Arg Arg
210 215 220

Met Pro Asp Gln Tyr Asp Arg Thr Leu Arg Glu Ile Phe Pro Asp Gln
225 230 235 240

His Pro Gly Gly Phe Ser Gln Leu Glu Asp Gly Arg Trp Val Trp Thr
245 250 255

Thr Phe Asn Ser Phe Gln Trp Asp Leu Asn Tyr Ser Asn Pro Trp Val
260 265 270

Phe Arg Ala Met Ala Gly Glu Met Leu Phe Leu Ala Asn Leu Gly Val
275 280 285

Asp Ile Leu Arg Met Asp Ala Val Ala Phe Ile Trp Lys Gln Met Gly
290 295 300

Thr Ser Cys Glu Asn Leu Pro Gln Ala His Ala Leu Ile Arg Ala Phe
305 310 315 320

Asn Ala Val Met Arg Ile Ala Ala Pro Ala Val Phe Phe Lys Ser Glu
325 330 335

Ala Ile Val His Pro Asp Gln Val Val Gln Tyr Ile Gly Gln Asp Glu
340 345 350

Cys Gln Ile Gly Tyr Asn Pro Leu Gln Met Ala Leu Leu Trp Asn Thr
355 360 365

Leu Ala Thr Arg Glu Val Asn Leu Leu His Gln Ala Leu Thr Tyr Arg
370 375 380

His Asn Leu Pro Glu His Thr Ala Trp Val Asn Tyr Val Arg Ser His
385 390 395 400

Asp Asp Ile Gly Trp Thr Phe Ala Asp Glu Asp Ala Ala Tyr Leu Gly
405 410 415

Ile Ser Gly Tyr Asp His Arg Gln Phe Leu Asn Arg Phe Phe Val Asn
420 425 430

Arg Phe Asp Gly Ser Phe Ala Arg Gly Val Pro Phe Gln Tyr Asn Pro
435 440 445

Ser Thr Gly Asp Cys Arg Val Ser Gly Thr Ala Ala Ala Leu Val Gly
450 455 460

Leu Ala Gln Asp Asp Pro His Ala Val Asp Arg Ile Lys Leu Leu Tyr
465 470 475 480

Ser Ile Ala Leu Ser Thr Gly Gly Leu Pro Leu Ile Tyr Leu Gly Asp
485 490 495

Glu Val Gly Thr Leu Asn Asp Asp Asp Trp Ser Gln Asp Ser Asn Lys
500 505 510

Ser Asp Asp Ser Arg Trp Ala His Arg Pro Arg Tyr Asn Glu Ala Leu
515 520 525

Tyr Ala Gln Arg Asn Asp Pro Ser Thr Ala Ala Gly Gln Ile Tyr Gln
530 535 540

Gly Leu Arg His Met Ile Ala Val Arg Gln Ser Asn Pro Arg Phe Asp
545 550 555 560

Gly Gly Arg Leu Val Thr Phe Asn Thr Asn Asn Lys His Ile Ile Gly
565 570 575

Tyr Ile Arg Asn Asn Ala Leu Leu Ala Phe Gly Asn Phe Ser Glu Tyr
580 585 590

Pro Gln Thr Val Thr Ala His Thr Leu Gln Ala Met Pro Phe Lys Ala
595 600 605

His Asp Leu Ile Gly Gly Lys Thr Val Ser Leu Asn Gln Asp Leu Thr
610 615 620

Leu Gln Pro Tyr Gln Val Met Trp Leu Glu Ile Ala
625 630 635

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Neisseria polysaccharea

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CTCACCATGG GCATCTTGGA CATC

24

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Neisseria polysaccharea

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CTGCCATGGT TCAGACGGCA TTTGG

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/EP 98/05573

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/54 C12N9/10 C12N1/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 95 31553 A (INST GENBIOLOGISCHE FORSCHUNG ;KOSSMANN JENS (DE); BUETTCHER VOLKE) 23 November 1995 see claims	1-12
X	DE 44 20 223 C (INST GENBIOLOGISCHE FORSCHUNG) 4 May 1995 see claims	1-12

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 May 1999	18/05/1999
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Delanghe, L

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. .ional Application No
PCT/EP 98/05573

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9531553 A	23-11-1995	DE	4417879 A	23-11-1995
		DE	4447388 A	27-06-1996
		AU	699552 B	10-12-1998
		AU	2614195 A	05-12-1995
		CA	2190149 A	23-11-1995
		EP	0759993 A	05-03-1997
		HU	76087 A	30-06-1997
		JP	10500297 T	13-01-1998
DE 4420223 C	04-05-1995	AU	696978 B	24-09-1998
		AU	2787895 A	04-01-1996
		CZ	9603603 A	16-04-1997
		WO	9533838 A	14-12-1995
		EP	0760856 A	12-03-1997
		HU	76348 A	28-08-1997
		JP	10504182 T	28-04-1998